

Yiliu YANG

(765) 421-1120 | yyiliu@cmu.edu | Pittsburgh, PA
yyiliu.com | [linkedin.com/in/yyiliu/](https://www.linkedin.com/in/yyiliu/)

EDUCATION

Carnegie Mellon University <i>Master of Science in Intelligent Information Systems (MIIS)</i>	Pittsburgh, PA, USA Expected Dec 2027
The Chinese University of Hong Kong <i>Bachelor of Engineering in Artificial Intelligence</i> GPA: 3.82/4.00, Major GPA: 3.92/4.00, Rank: 1 st of 39 Honors: First Class Honours, ELITE Stream, Dean’s List (2020-2024) Graduate-level Coursework: Machine Learning Theory, Foundations of Optimization, Pattern Recognition	Hong Kong Jul 2024

SKILLS

Programming: Python, C, C++, SQL
LLM/ML: LLM Agents, PyTorch, RAG, vLLM, Transformers, LoRA
Engineering: Linux, Git, Docker, Redis, FastAPI
Languages: English (TOEFL 105), Mandarin (Native)

PROFESSIONAL EXPERIENCE

RedotPay <i>AI Engineer</i>	Hong Kong Mar 2026 – Aug 2026
- Co-developed and productionized an AI customer-service agent serving approximately 3,500 daily conversations with knowledge-grounded responses, MCP tools, SOP workflows, conversation analytics, and human handoff - Compared with the legacy Dify bot, the agent increased CSAT from 71.6% to 84.0%, AI-only resolution rate from 79.2% to 85.6%, and deflection rate from 70.6% to 81.9%, saving an estimated 50 support hours per day - Designed and built a 909-case automated regression framework for tool-using workflows, validating required and prohibited MCP calls with mocked backend states across KYC and SOP scenarios	

RESEARCH EXPERIENCE

Multimedia Laboratory, The Chinese University of Hong Kong <i>Volunteer Research Assistant</i>	Hong Kong Mar 2025 – Feb 2026
- Developed <i>QuadSentinel</i>, a multi-agent guard system translating natural language policies into machine-checkable rules for robust runtime safety enforcement - Achieved 91.5% accuracy on the AgentHarm safety benchmark across prompt injection, jailbreak, and tool misuse attacks, outperforming prior state-of-the-art methods	

PUBLICATIONS

QuadSentinel: Sequent Safety for Machine-Checkable Control in Multi-agent Systems, Yiliu Yang*, Yilei Jiang*, Qunzhong Wang, et al., arXiv:2512.16279
LogicAsker: Evaluating and Improving Logical Reasoning Ability of Large Language Models, Yuxuan Wan*, Wenxuan Wang*, Yiliu Yang, et al., EMNLP 2024

LEADERSHIP

International Collegiate Programming Contest (ICPC) <i>CUHK Team Leader</i>	Hong Kong Oct 2020 – Aug 2023
Led a three-member competitive programming team to four regional bronze medals in Macau, Kunming, Shenyang, and Hong Kong	

SELECTED HONORS AND AWARDS

Gold Award for Outstanding Academic Performance	2023, 2024
HKSAR Government Scholarship	2023